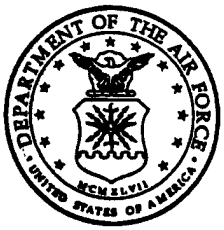
CTN Test Report 91-028

MODEL--TECHNICAL MANUAL DATA



31 March 1992

Prepared for

Air Force Materiel Command

CTN Test Report
91-028

31 March 1992

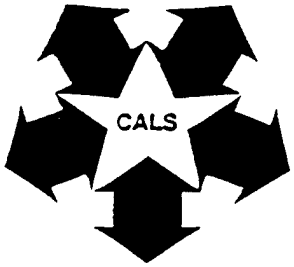
FINAL

Model -- Technical Manual Data

Prepared for:
Air Force Materiel Command
CALS Test Network (CTN) (AFMC/ENCT)
Wright-Patterson AFB, OH 45433-5000

19 April 1991

PEO STAMIS



By:

Department of the Army
PM CALS

CTN Contact
Mel Lammers
(513) 257-3085

FINAL

Model -- Technical Manual Data

CONTRACT NO. DAAB07-89-D-A047
TASK ASSIGNMENT PLAN NO. 90-006

Prepared for:

Department of the Army
PM CALS
Ft Monmouth NJ 07703

19 April 1991

Army CALS Test Bed Contact
Alton K. Fairweather
(908) 532-0414

By:

ACCURATE
Information Systems, Inc.

DISCLAIMER

This report and those involved in its preparation do not endorse any product, process, or company stated herein. Use of these means by anyone does not imply certification by the CALS Test Network.

CONTENTS

LIST OF FIGURES

EXECUTIVE SUMMARY

The Department of Defense (DoD) service organizations have a huge store of technical publications that are required to support weapon systems. The publications are predominantly paper-based and create copious management, storage and distribution problems. The DoD's Computer-aided Acquisition and Logistic Support (CALS) strategy provides for transition from the paper-intensive life cycle processes of these publications to the utilization of digital information technologies. The acceptance of digital data for storage in a repository poses new problems that are being addressed by the CALS Test Network (CTN) in cooperation with the service organizations. Data Acceptance (DA) addresses the quality of the final deliverable CALS compliant digital data files before they are stored in the repository. Data Acceptance (DA) addresses the quality of the deliverable files, not the format or technical accuracy of the document itself.

This document presents models that define the entities and attributes related to the acceptance of CALS compliant digital data for technical manuals. The models are expandable as the CALS standards mature. An entity relationship methodology is used to define a data model. The data model depicts the relationships of data types at various levels of detail. The procurement of CALS technical manual data are depicted in process models which show the generation and the acceptance of the technical publication CALS data. Digital data quality criteria are applied to the acceptance or rejection of the data for each type of document: Raster Page Image, Standard Generalized Markup Language (SGML) tagged text, or other.

Raster Page Image documents are produced by converting from hardcopy or microfiche format to a raster page image format. This model emphasizes the verification of the file image data quality and file header accuracy. The SGML Document acceptance function is expanded in a model that depicts the verification of the file header, text tagging, and illustration file quality.

This document concludes with a discussion of the contract issues that should be considered in the acceptance of technical manual digital data. Specific recommendations are included that will enhance the quality of digital data for technical manuals generated by contractors and provide guidelines for the acceptance of that data by the government.

1.0 INTRODUCTION

The Department of Defense (DoD) armed services, Air Force, Navy, Army, etc., have a vast store of paper-based technical publications that are required to support the weapon systems in the field.

The Computer-aided Acquisition and Logistic Support (CALS) strategy provides for transition from the paper-intensive weapon system life cycle processes to the use of digital information technology. Acquisition of new weapon system technical publications will be in digital data form. Existing active weapon systems also may be required to convert paper-based supporting technical publications to digital form. The result will be more efficient storage, access and distribution of the technical publication data.

The converted Technical Manual (TM) data and the newly acquired TM data must comply with the CALS standards. These include MIL-D-28000, MIL-M-28001, MIL-R-28002A, MIL-D-28003, and MIL-STD-1840A. CALS conformance requires close cooperation between industry and government because there are a variety of technical publications systems and authoring systems with proprietary formats. Conversion to and from CALS formats will be necessary in both authoring systems and publishing systems, respectively.

Existing procurement of TMs is by camera-ready hardcopy, microfiche, and digital data. The quality of hardcopy and microfiche deliverables is defined in terms of legibility and reproducibility, which have an impact on all copies reproduced from the deliverables. At present, the government users are proficient in the acceptance of paper-based TM products from the contractors. Although acceptance of the TM is successful, the process is labor-intensive.

The skills for accepting quality paper-based TM products must now be transferred to the acceptance of digital data. The quality of digital data is defined in terms of specification compliance and processing. This involves adapting the visual inspection process from paper or film-based display to the electronic display. However, this is also labor-intensive. Computer-assisted techniques can be used to greatly reduce the amount of data which must be visually inspected for quality. Data Acceptance will verify the quality of the digital data prior to permanent storage in the government repository.

A model is presented which will form the basis for the development of procedures for the acceptance of the CALS compliant digital data for Technical Manuals. As the CALS standards mature, the model can be revised accordingly. For this

model, it is assumed that the acquisition, creation, contractor validation and government verification of the technical content, completeness and accuracy of the TM are done separately and are independent of Data Acceptance.

2.0 PURPOSE

This document presents a model that addresses those entities and attributes associated with the acceptance of CALS compliant digital data for Technical Manuals.

The model describes, at a high level, those functions required to accept TM text and graphics data when procured as raster, Standard Generalized Markup Language (SGML) mark-up or a combination where illustrations are in raster, Initial Graphics Exchange Specification (IGES), or Computer Graphics Metafile (CGM). The model can be used for the development of manual and computer-assisted data acceptance procedures for use by member agencies of the DoD.

3.0 SCOPE

The model describes the major activities required to perform data acceptance of CALS compliant digital data for TMs. The methodology used to develop the process and data models depicted are described. Technical Manual data entities are defined with references to the associated CALS standards. A brief overview of existing procurement of technical manual data is presented and CALS procurement requirements are outlined. Procurement of CALS compliant digital data involves new TMs or older TMs that will be selected for conversion from hardcopy, microfiche, or proprietary digital data. The main section addresses data acceptance of the TM component data types. A concluding section describes contract considerations that are key to the procurement of CALS compliant TM digital data.

The data acceptance described by the model is generic in that it is independent of hardware, software, contractors, government agencies, etc. It is applicable to a variety of procurement scenarios and a variety of hardware and software platforms.

4.0 APPROACH

The approach taken in development of this model required analysis of existing procedures for the acceptance of technical manual data in hard copy format. The user community within the tri-services have developed expertise in the QA of the TM product. Inputs were obtained from the tri-services in an effort to know what procedures are presently being used. The next effort was to relate the existing procedures and expertise to the acceptance of digital technical manual data, taking into consideration the requirements for CALS compliance.

A methodology was defined for presenting this information in the form of process flow diagrams and entity relationship diagrams. The model was developed using this methodology to define the process for the generation of the technical publication data at the contractor site and the acceptance of the digital technical publication data by the government.

This model also considers the impact that the procuring document, the contract, has in the acceptance of the data. The ultimate quality of the data is greatly dependent on the source development, data validation/verification and future support as defined in the contractor's data certification and warranties.

The model attempts to address those entities and attributes most directly associated with the acceptance of technical manual data as well as other key issues that may indirectly affect the quality of the technical manual data.

5.0 METHODOLOGY

The model is documented with the use of process flow diagrams and entity relationship diagrams. A CASE tool was used to assist in preparing the model for presentation. The Gane & Sarson methodology was used to model the process flow. The entity relationship methodology was used to model the digital data.

5.1 Process Modeling

The process flow is illustrated as a hierarchy of Gane & Sarson process flow diagrams starting with an overview process flow. Each process flow diagram describes a breakdown of an overall procedure. It shows data stores, processes that operate on them, and the data flow which supports the processing. The symbols used in the process flow diagrams are shown in Figure 1.

A process is represented by a rectangular figure with rounded corners and labeled with the letter "P." A process denotes a distinct procedure, step, or other breakdown of a larger process. A data store is a rectangular figure that is open on the right-hand side and labeled with the letter "D." Data flow is represented as a directed line segment that is connected to the origin of the data and points to the destination of the data. A data flow implies that the data exists temporarily while a data store implies some permanence.

Data stores in a diagram do not specify the storage medium; data may be digital data, paper documents, verbal instructions, common knowledge, etc. Data flows shown in a diagram do not specify the transfer medium; data transfer may be via magnetic media, telecommunications, mail, spoken word, etc. Similarly, processes shown in a diagram do not specify the means by which they are performed; processing may be computer processing, machine processing, manual processing, etc. A process flow diagram illustrates all data stores and processes, and all possible data flows. For a given scenario, a subset of the all processes and data may be applicable. The diagram does not show a sequence of events since the sequence may vary for each possible scenario.

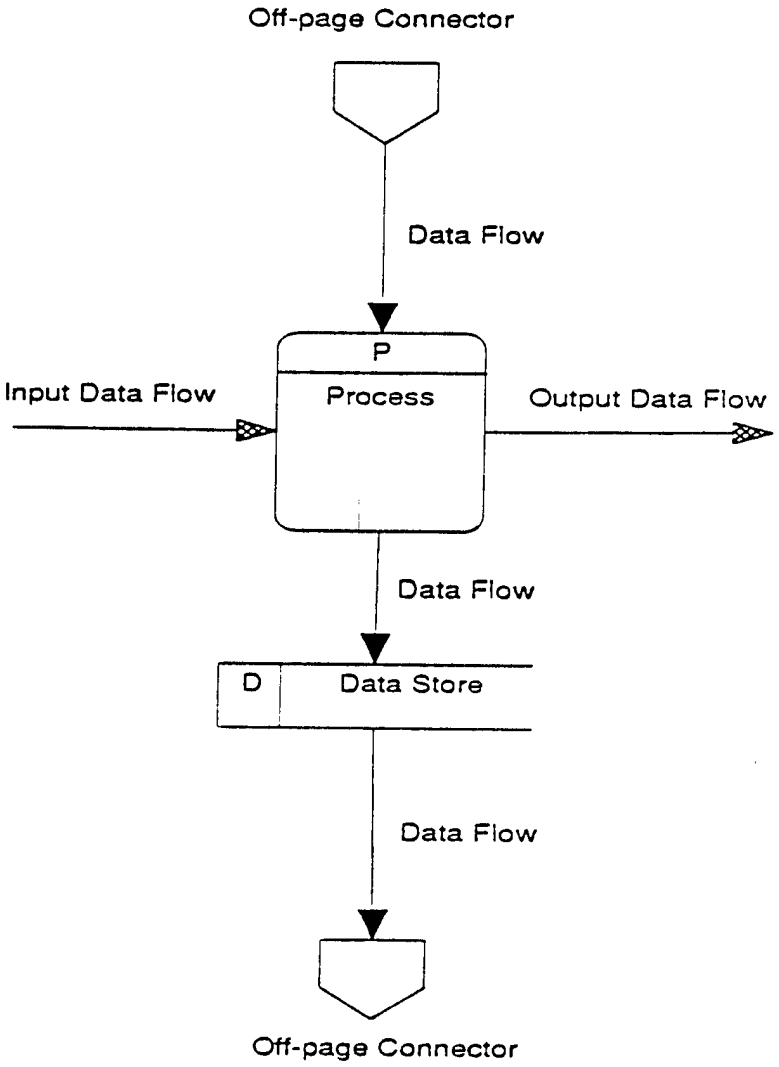


Figure 1 - Gane & Sarson Methodology Symbol Legend

5.2 Data Modeling

The data structure is illustrated as Entity Relationship (ER) diagrams. Each ER diagram shows the relationship among various types of data, or entities. The symbols used in the process flow diagrams are shown in Figure 2.

Each entity is represented by a rectangle with rounded corners. The relationship of an entity with another entity, or the relation, is shown as a line connecting the two entities. The type of relation is shown graphically by the connection to each of its entities. The following are types of relations that are used in the data model.

zero or one	This entity is optional. There can be, at most, one instance of it.
zero or many	This entity is optional. There can be one or more instances of it.
one and only one	This entity is required. There can only be a single instance of this entity.
one or more	This entity is required. There must be at least one instance of it. There can be many instances of it.
many	This entity is required. There must be at least two instances of it. There can be many instances of it.

At the bottom of Figure 2, a simple example illustrates the relation between Raster Page Image Documents and Raster Page Image Files: for each Raster Page Image Document, there is at least one, possibly many, Raster Page Image File(s). Literally, it states that "for one and only one Raster Page Image Document, there are one or more Raster Page Image Files."

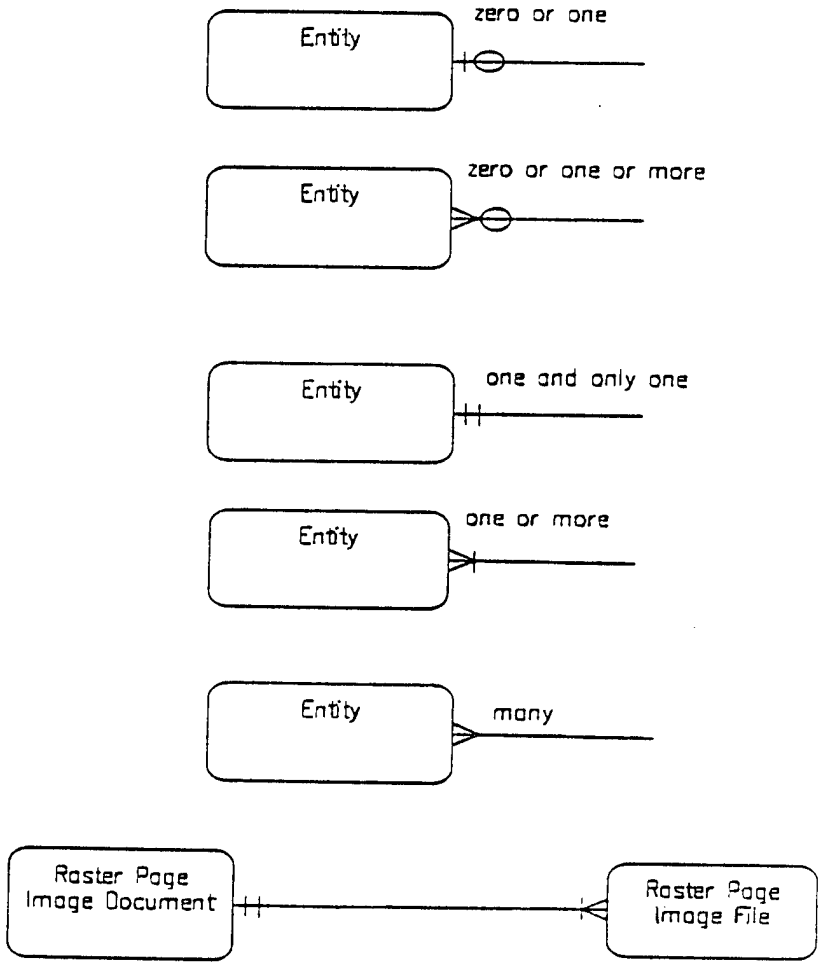


Figure 2 - Entity Relationship Methodology Symbol Legend

6.0 DATA MODEL

Technical Manual Data consists of many interrelated component types of data. This section defines the data and its relationships in order to develop a usable model of data. This is done to clarify which data types are affected by various procurement issues and data acceptance issues.

Only CALS compliant data is being considered in the model. That is, only data exchange is considered. Proprietary data that is resident in an authoring system, in a publishing system, or in a government agency repository is not being considered in the data model.

The relationship among the data types is shown graphically in an entity relationship diagram in Figure 3. The following are definitions of Technical Manual data types and are used in the discussion of the procurement process in this document. Definitions are taken from MIL-STD-1840A, MIL-D-28000, MIL-M-28001A, MIL-R-28002A, and MIL-D-28003.

File Set. A File Set is a set of data files and their associated declaration files, together which make up a single Document.

Technical Publication. A Technical Publication is a single Document which can be referred to as either a Technical Manual or a Technical Order.

Technical Manual. A Technical Manual is a term for a Technical Publication used by the Army and Navy.

Technical Order. A Technical Order is a term for a Technical Publication used by the Air Force.

Document. A Document is a Technical Publication which is stored in a File Set. A document can be either a Raster Page Image Document, an SGML Document, or Other (type of) Document.

Raster Page Image Document. A Raster Page Image Document is a type of Document which is composed of one or more Raster Page Image Files. Raster Page Image Documents are described in MIL-STD-1840A Section 4.1.1.1.1.

SGML Document. An SGML document is a type of Document which is a document instance composed of one or more text data files, any number of optional illustration files, a DTD file, and a FOSI file. An SGML Document is

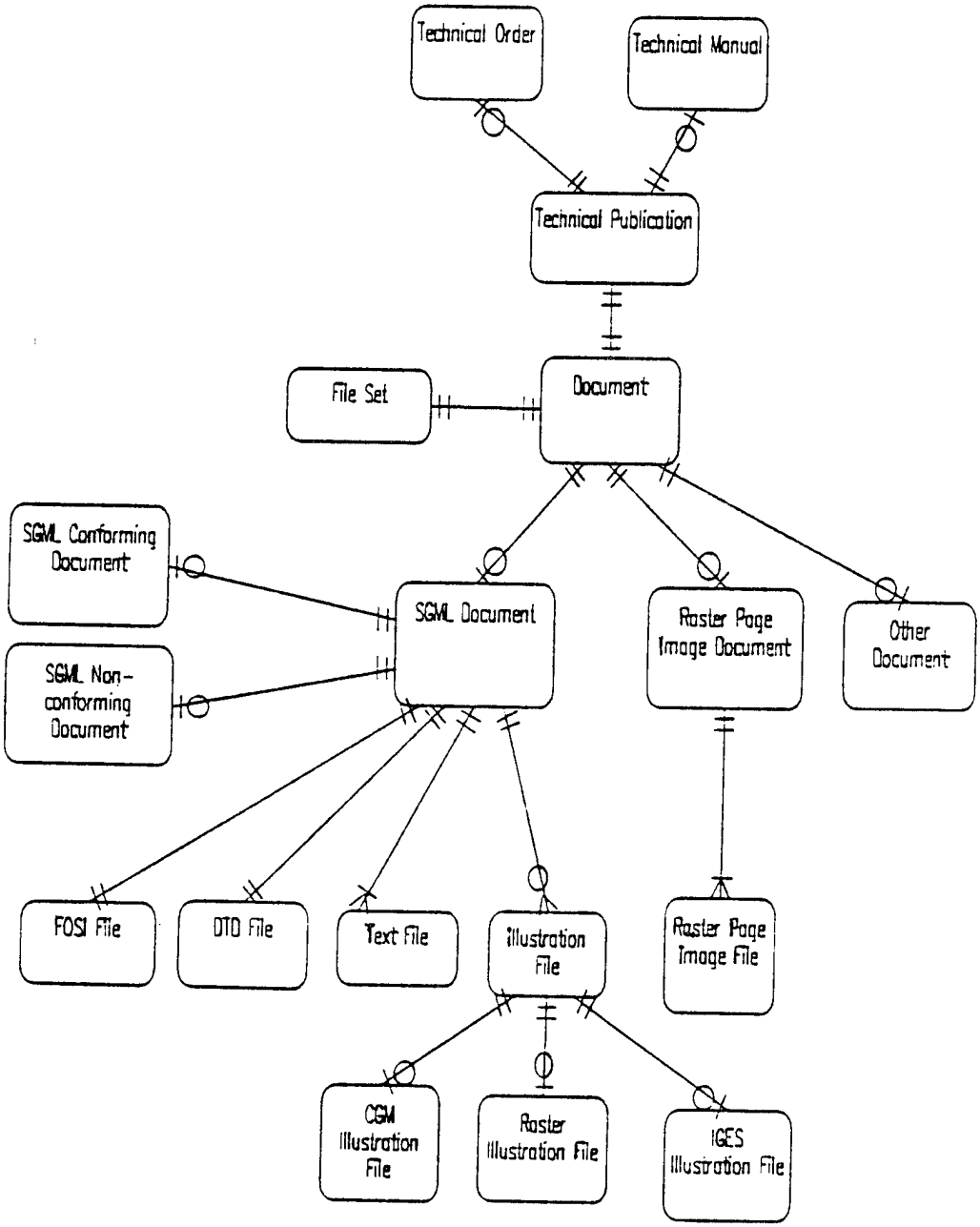


Figure 3 - Data Model

described in MIL-M-28001. An SGML Document can be either a Conforming SGML Document or a Non-conforming SGML Document.

Conforming SGML Document. A Conforming SGML document is a type of SGML Document whose SGML tags conform to the baseline tag set defined in MIL-M-28001. Conforming documents are not delivered with DTD or FOSI files. Conforming SGML Documents are described in MIL-STD-1840A Section 4.1.1.1.3.

Non-conforming SGML Document. A Nonconforming SGML Document is a type of SGML Document whose SGML tags do not conform to the baseline tag set defined in MIL-M-28001. Nonconforming documents are delivered with DTD and FOSI files. Conforming SGML Documents are described in MIL-STD-1840A Section 4.1.1.1.4.

Other Document. An Other (type of) Document can be a Page Description Language (PDL) Document or other type of document to be described by MIL-STD-1840A in the future. PDL Documents are described in MIL-STD-1840A Section 4.1.1.1.2.

Raster Page Image File. A Raster Page Image File contains data which is a pixel image of a scanned document page that is encoded using CCITT Group 4 compression as specified in MIL-R-28002A. Raster Page Image Files are described in MIL-STD-1840A Section 4.1.1.2.9.

Illustration File. An Illustration File is an optional component of a Conforming SGML Document or a Nonconforming SGML Document. An Illustration File consists of a graphical representation contained within a technical publication. Illustration Data are encoded in IGES, raster, or CGM format. Illustration Files are defined in MIL-STD-1840A Section 4.1.1.2.5.

Raster Illustration File. A Raster Illustration File is a type of Illustration File. A Raster Illustration File contains a pixel image of a document illustration encoded using CCITT Group 4 compression as specified in MIL-R-28002A. Raster Illustration Files are described in MIL-STD-1840A Section 4.1.1.2.5.1.

IGES Illustration File. An IGES Illustration File is a type of Illustration File. An IGES Illustration File contains a vector representation of an image of a document illustration encoded using the Initial Graphic Exchange Specification (IGES) format as specified in MIL-D-28000. IGES Illustration Files are described in MIL-STD-1840A Section 4.1.1.2.5.1.

CGM Illustration File. A CGM Illustration File is a type of Illustration File. A CGM Illustration File contains a vector representation of an image of a document illustration encoded using the Computer Graphics Metafile (CGM) format as specified in MIL-D-28003. CGM Illustration Files are described in MIL-STD-1840A Section 4.1.1.2.5.3.

Text File. A Text File is a component of a Conforming SGML Document or a Nonconforming SGML Document. A Text File contains text data coded as ASCII characters and tagged with SGML tags which are defined in the baseline tag set defined in MIL-D-28001. A Text File is described in MIL-STD-1840A Section 4.1.1.2.2.

DTD File. A Document Type Definition (DTD) File is a component of a Conforming SGML Document or a Nonconforming SGML Document. A DTD File is an ASCII text file which contains SGML tag definitions. DTD Files are described in MIL-STD-1840A Section 4.1.1.2.3.

FOSI File. A Formatting Output Specification Instance (FOSI) File is a component of a Conforming SGML Document or a Nonconforming SGML Document. A FOSI File is an ASCII text file which contains formatting definitions for SGML tags. FOSI Files are described in MIL-STD-1840A Section 4.1.1.2.4.

7.0 DIGITAL DATA QUALITY CRITERIA

Digital data quality is defined by specific criteria which are applied by the government agency to data to determine whether it can be accepted or must be rejected. Some data quality criteria apply to all types of data. Other data quality criteria are different for each type of document: raster page image, SGML conforming, SGML non-conforming, or other.

It is assumed that the publication content has been reviewed and approved. Digital data quality criteria do not address the content at all.

7.1 Data Format

The format of the deliverable files must comply with MIL-STD-1840A specifications. Files must have specified header records, and deliverables must have specified declaration files.

The declaration file format (size and record order by record identifier) must be in accordance with MIL-STD-1840A. The file counts in the 'filcnt:' record of the declaration file must agree with the actual file counts in the deliverable file set.

The data file format (size and record order by record identifier) must be in accordance with MIL-STD-1840A. The data file content format must agree with the file name code, the fifth character in the file name. Text files, whose name code is 'T', are checked for compliance with the standards defined in MIL-M-28001. DTD files, whose name code is 'G', are checked for compliance with the standards defined in MIL-M-28001. FOSI files, whose name code is 'H', are checked for compliance with the standards defined in MIL-M-28001. Raster image files, whose name code is 'R', are checked for compliance with the standards defined in MIL-R-28002A. IGES files, whose name code is 'Q', are checked for compliance with the IGES standards defined in MIL-D-28000. CGM files, whose name code is 'C', are checked for compliance with the CGM standards defined in MIL-D-28003.

The criteria which can be tested depend on the environment in which the files are verified. If the files are verified on a deliverable physical medium, the format of the files and their counts can be tested. If the files are verified before they are copied on a deliverable medium, only the format of the individual files can be tested.

7.2 Data Validity

The validity of the data in the deliverable files must be in accordance with its intended use. This requires compliance with standards or practices that may be specific to each agency.

For example, raster illustration data must have a resolution of 300 dots per inch or greater. This can be verified by examining the 'rpelcnt:' header record of raster data files. If necessary, it can further be verified by examining the image data itself.

Vector illustration data must make efficient use of standard entities. For example, to define a circle, it is valid to use a circle entity, not a large number of line segment entities which approximate the circle. A count of entity types can indicate unusual deviations from those of normal drawings.

7.3 Image Quality

Image quality is a characteristic of raster images. It is the degree of legibility and reproducibility of the image. Three key image quality criteria are:

Density: Image quality is a function of density, or the amount of information contained in the image. Image density can vary but it should be consistent for groups of similar images. Poor quality images will appear significantly darker or lighter than the acceptable range for their type.

Noise: Image noise appears as black and white orphan pixels superimposed on a raster image. An orphan is a pixel or a small group of pixels that is completely surrounded by the contrasting color. A black orphan is a dark spot surrounded by white space. A white orphan is a white speckle in a filled-in image area (e.g., a line, a character, etc.). An orphan pixel is likely to represent noise instead of image data. An excess of orphan pixels is likely to be noise introduced by the image generation process.

Verticality: Verticality is the angle of orientation of the image with respect to the viewer's reference frame. An excessively skewed image is likely to be missing information due to cropping at the corners.

Image quality is an issue primarily when evaluating raster data that was scanned and digitized from hard copy or microfiche sources. The legibility of the raster image may be poor because the source image legibility was poor or because noise was introduced by the digitizing process. The image may be skewed because the source image on microfiche was skewed or because the vertical source image was skewed when it was scanned and digitized.

Raster image quality is an issue for entire raster page image documents and illustration files of SGML documents. Image quality is not as important an issue when dealing with raster image data files that were generated directly from character data or vector illustration data.

A related issue of some importance is the degree of compressibility of raster images. Dark images and noisy images both have an excess number of dark pixels which are stored as information by the Group 4 compression. This reduces their compressibility and increases the storage requirements.

7.4 Identification Data

Identification data is indexing information which is stored in data file headers and in declaration file records. This information is used by a government agency's storage, retrieval, and publishing system.

Identification data quality depends on the accuracy of the data in a file header. It must agree with the data in the file and the data in the declaration file. Poor identification data quality results when there is a discrepancy among any of these.

Identification data quality checking is performed by obtaining the identification data from the contents of the file and comparing it with that in the file header. If the identification data cannot be obtained from the data file, this also constitutes poor identification data quality.

In all documents, the document number in the 'dstdocid:' record of each file must match the document number in the 'dstdocid:' record of the declaration file.

In raster page image documents, there must be an exact match between the page number data in the 'txtfilid:' record of the image file header and the page number in the digitally stored page image. The page number identification data can be obtained from the page image file by applying character recognition techniques to convert pixel data in the margin area to ASCII text.

In SGML documents, there must be an exact match between the graphic identifier in the 'srcgph:' record of the file header and the 'boardno' attribute of each <graphic ...> entity which uses it. Illustration files which have incorrect 'srcgph:' record data can result in one or more of three anomalies: missing figures, unused figures, and incorrect figures. Missing and unused figures can be detected automatically by an SGML parser which verifies the existence of each file which is referenced by the boardno attribute of the graphic entity and maintains a count of references for each file. Incorrect figures can only be detected by visual inspection of the formatted output.

In SGML documents, there must be an exact match between the document content code in the 'txtfilid:' record of the illustration files and the document content code in the 'txtfilid:' record of the text file which references them. Mismatched document content codes can be detected automatically by an SGML parser which verifies that the code of each illustration file specified in a <graphic ...> entity matches the code of the text file which references it.

In SGML documents, the figure identifier in the 'figid:' record of the illustration files must be correctly assigned. In documents which have figure numbers, there must be an exact match between the figure identifier in the 'figid:' record of the illustration files and the figure identifier generated automatically when creating output using the appropriate DTD and FOSI. Mismatched figure identifiers can be detected automatically by an SGML parser which verifies that the figure identifier of each illustration file specified in a <figure ...> entity matches the automatically generated figure identifier. In documents which do not have figure numbers, the figure identifier in the 'figid:' record must be in accordance with a naming or numbering scheme required by the contract or other form of agreement.

7.5 Character Recognition Accuracy

Character recognition accuracy is an issue in the conversion of existing manuals to text character data using character recognition techniques. The resulting text character data must be identical in content to the original hardcopy or microfiche document. Discrepancies can arise from unrecognized or incorrectly recognized text. This may be due to poor scanned image quality or from limitations in the character recognition algorithm(s) used.

Text accuracy can initially be verified by subjecting the generated text to spell checking and grammar checking. When recognized letters and numbers are missing or incorrect, the resulting spelling and grammar usually will be grossly incorrect. This technique will uncover many character recognition failures or mistakes and can identify problem areas in character recognition.

7.6 SGML Syntax

Tagging in the document instance is an issue in both the creation of new manuals and in the conversion of existing manuals to character data. Only tags defined in the baseline tag set defined in MIL-M-28001 are allowed. The document instance must be tagged according to the rules defined in the appropriate DTD. The document instance must be tagged to the appropriate depth as defined in the contract or other form of agreement. SGML syntax can be checked by an SGML parser.

7.7 DTD Syntax

DTD syntax is an issue in both the creation of new manuals and in the conversion of existing manuals to character data. The DTD must have valid definitions of SGML tags. SGML tags used must be among those defined in Appendix A of MIL-M-28001A. DTD syntax can be checked by an SGML parser.

7.8 FOSI Syntax

FOSI syntax is an issue in both the creation of new manuals and in the conversion of existing manuals to character data. The FOSI must be written in accordance with the output specification in MIL-M-28001. FOSI syntax can be checked by an SGML parser.

8.0 PROCUREMENT

This section gives a brief overview of the procurement of CALS compliant technical manual digital data for storage in and retrieval from a repository. This can be either new technical manual procurement or the conversion of existing technical manuals to CALS compliant digital data.

8.1 CALS Compliance

A government agency usually must convert digital data from contractor supplied formats to a native format. A lack of compliance with format standards and/or a lack of documentation of exchange formats is a problem faced by government agencies. Consequently, a government agency must spend considerable time to examine data formats supplied and to convert the data into a useable format.

CALS compliance will be required for delivery of digital data to the government agency. Exchange of CALS compliant data will alleviate the problem of format conversion because there will be published format standards which can be used as contract requirements. This will improve the quality of the data delivered to the government agency because it will conform to published standards. It will also reduce the government agency time spent on converting the data by placing responsibility on the contractor to deliver the data in the contract specified published formats.

CALS compliance is required for the exchange of digital data. Authoring systems and publishing systems must be able to produce and accept, respectively, CALS compliant data. An authoring or publishing system may have CALS compliant native formats or a system may have a translation function that performs CALS-to-native input translation and/or native-to-CALS output translation.

Translation from a contractor's publishing system native formats to CALS compliant formats will be an issue which must be resolved by contractor validation. The government agency will verify the contractor's ability to deliver CALS compliant data.

Translation from CALS compliant formats to a government agency's publishing system native formats will be another issue. A translation facility must be available and it must be verified.

8.2 Procurement of New Manuals

The procurement of new manuals assumes that the technical manual content and format have been approved. This implies that the technical manual has been subjected to a validation walkthrough by the contractor, a verification walkthrough by the government agency, and a final review by the government agency. Therefore, the content of the technical manual is no longer an issue for data acceptance, whereas the format and quality of the final deliverable(s) are issues.

An issue in the acceptance of digital data is the ability of the government agency to produce the desired presentation format of the document. This may involve recreating the publication as it was originally created by the contractor. Alternatively, it may involve producing the output in a presentation format other than that supplied by the contractor.

8.3 Conversion of Existing Manuals

Existing manuals may be hardcopy, microfiche, or a proprietary form of digital data which is incompatible with the government agency's publishing system. Existing manuals can be converted to one of two forms of CALS compliant digital data: a Raster Page Image Document or an SGML text document with illustrations.

In a raster page image document, the document is scanned or converted one page at a time. Each scanned or converted page is stored in a file as a raster image and the file is indexed so that pages can be retrieved randomly. In an SGML text document, text on scanned pages is converted to character data using character recognition techniques and stored in text files. Illustrations are stored in separate files as raster data in CCITT Group 4 format or vector data in IGES or CGM format. They can be scanned from the original document or they can be re-created. Scanned illustrations can be stored directly as raster data or they can be converted by a vectorization process and stored as vector data.

The choice of conversion to raster page image data or tagged text data depends on economic considerations such as the cost of conversion versus the remaining service life of the information in the manual, the complexity of the manual, the frequency of revisions and updates, etc.

Tagged text and illustrations will be the easiest to index, retrieve, and update. However, when the cost of conversion does not justify the benefits, the manual can be converted to raster page images. This can be done for documents which are not expected to be updated or for documents which will be replaced in their entirety by an updated version.

In conversion of hardcopy and microfiche to raster page image data, the quality of the original images and the resulting quality of the digitally stored images are issues because copies will be produced directly from digitally stored images.

8.4 Procurement Overview

The procurement of Technical Manual Data is summarized in Figure 4.

Contract Award (P1). Procurement begins with a contract award which includes the procurement of technical manual data. The contract award is followed by Contractor Validation.

Contractor Validation (P2). The government agency verifies that the contractor is able to produce CALS compliant digital data. Contractor validation is required before the contractor can generate Technical Manual data.

Tech Pub Generation (P3). After contractor validation is completed, the contractor can generate the Technical Manual data which is sent to the government agency as a deliverable package. The digital data is delivered on the media specified in the contract. The media may be physical media, such as magnetic or optical, or telecommunications. For physical media, the media will be packaged in accordance with MIL-STD-1840A.

Data Acceptance (P4). The government agency can accept or reject the data. Accepted data is stored in a repository. The government agency should be able to return rejected data to the contractor for correction. The procurement should have provisions for the return and correction of digital data which is not accepted. The contract should specify what contract deliverables are returned and under what circumstances they are returned.

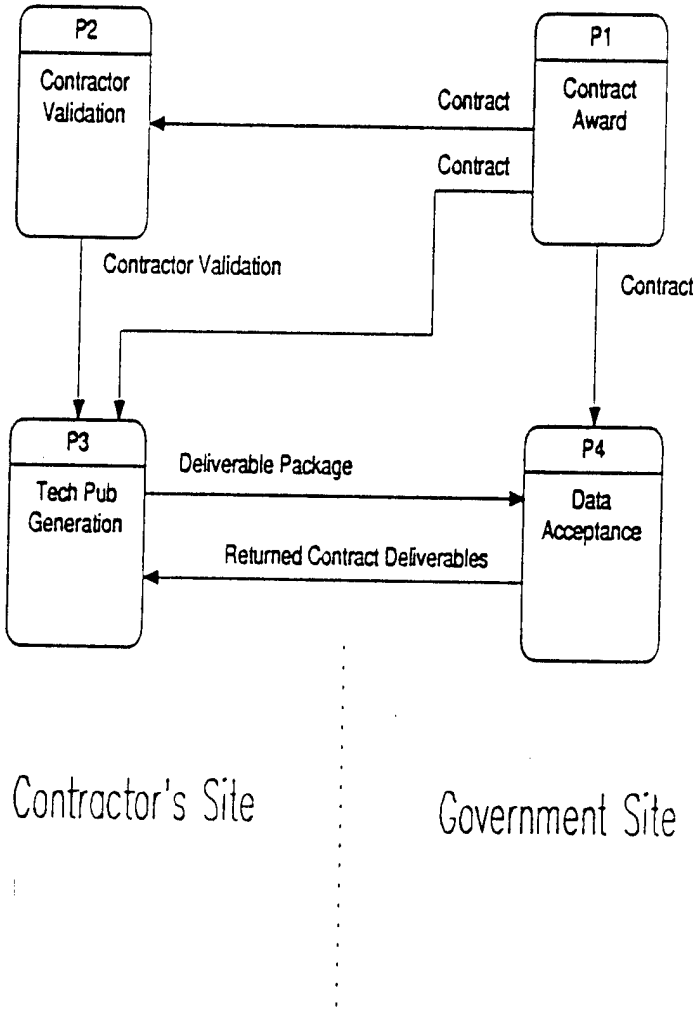


Figure 4 - Procurement

8.5 Tech Manual Generation

The generation of technical manual data occurs at the contractor site. Tech Manual (Pub) Generation is a subprocess (process P3) of the Procurement Overview in Figure 4. Technical Manual Generation is illustrated in Figure 5.

Generate Tech Pub Data (P3.1). Generation of a technical publication is the production of digital data for either new technical manual procurement or for conversion of existing manuals. The generation can proceed only after contractor validation is complete. For new manuals, this process will produce SGML tagged text documents with illustrations. For conversion of existing manuals it will produce raster page image documents or SGML tagged text documents with illustrations.

Translate Native to CALS Format (P3.2). After the Technical Publication data is generated, it must be translated from the contractor's native format to CALS format for delivery to the government agency. For physical media, this step involves copying the files on the media specified by the contract.

Generate Deliverable Package (P3.3). After the Technical Publication data is translated to CALS format, it is delivered to the government agency in accordance with the terms of the contract. Physical media are packaged and shipped in accordance with the specifications in MIL-STD-1840A. In the case of telecommunications, the files are transferred in accordance with the terms of the contract or other form of agreement.

Accept Returned Deliverables (P3.4). The procurement process should include provisions for the contractor to accept returned deliverables and correct problems identified in the data acceptance process.

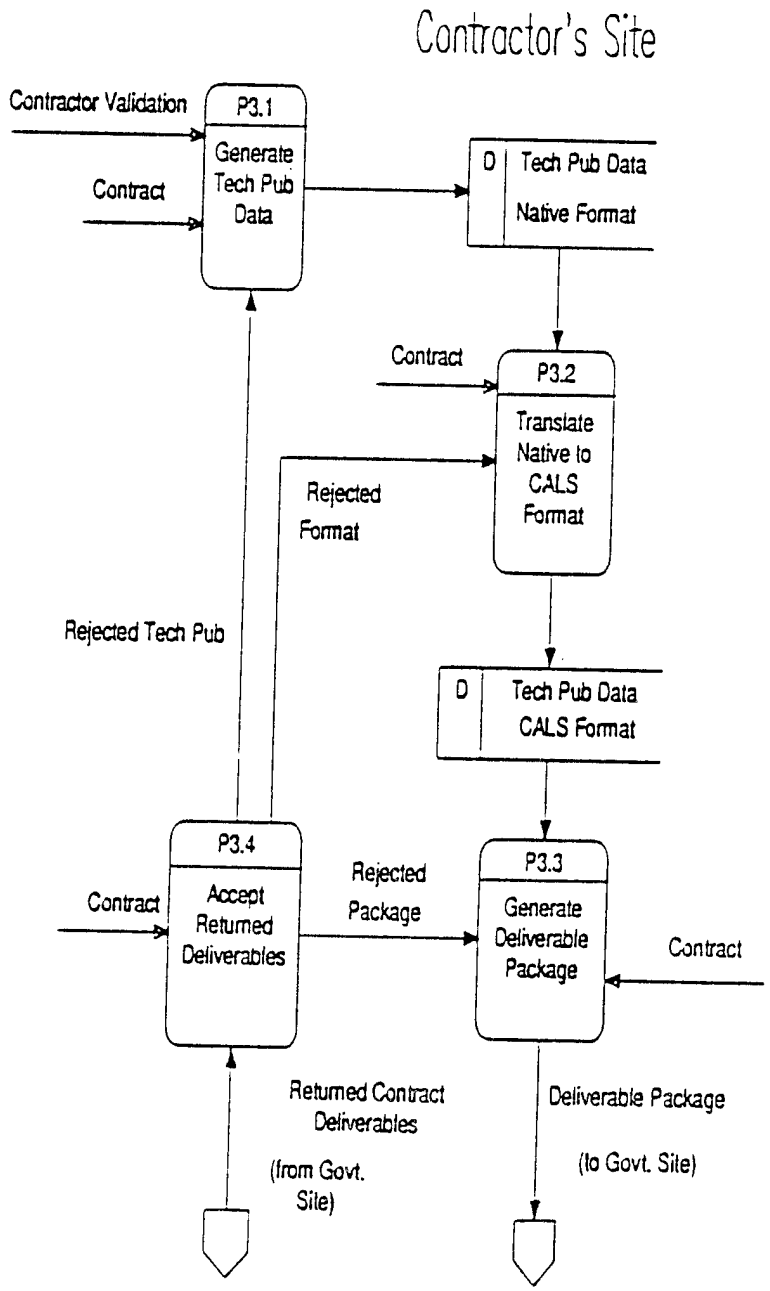


Figure 5 - Generation

8.6 Tech Manual Acceptance

After the digital data is generated and delivered by the contractor, it must undergo acceptance at the government agency. The government agency's Technical Manual team performs the acceptance. Tech Manual (Data) Acceptance is a subprocess (process P4) of the Procurement Overview in Figure 4. Tech Manual Acceptance is shown in Figure 6.

Accept Package (P4.1). First the deliverable package is inspected to verify that it contains deliverables specified in the contract and was packaged and shipped in accordance with MIL-STD-1840A. If the package fails inspection, it should be returned to the contractor in accordance with the contract on the basis of improper packaging and/or shipping. If the deliverable package is accepted, the digital data is loaded on a system where it undergoes data acceptance.

Data Acceptance (P4.2). Data Acceptance is the acceptance of digital data based on its conformance to CALS data exchange formats. Data Acceptance is a key function in the overall Technical Manual Acceptance process and is described in further detail in Section 9.0 on page 28.

Translate CALS to Native Format (P4.3). If the data is accepted by the data acceptance process, it is translated to the native format of the government agency's publishing system. The translated files are stored temporarily for final acceptance before they are stored permanently in the repository.

Final Acceptance (P4.4). After data is translated to the government agency's native format, it is subjected to Final Acceptance. Final Acceptance addresses the approval of the presentation format of the document based on review of output produced from the translated data. The formatted output can be hardcopy or it can be previewed on a screen. The output may be compared to an original document produced by the contractor and approved by the government agency. Alternatively, the output may be reviewed against other criteria as specified by the contract or other form of agreement. If the output is not acceptable based on the review criteria, the data is rejected and returned to the contractor for correction.

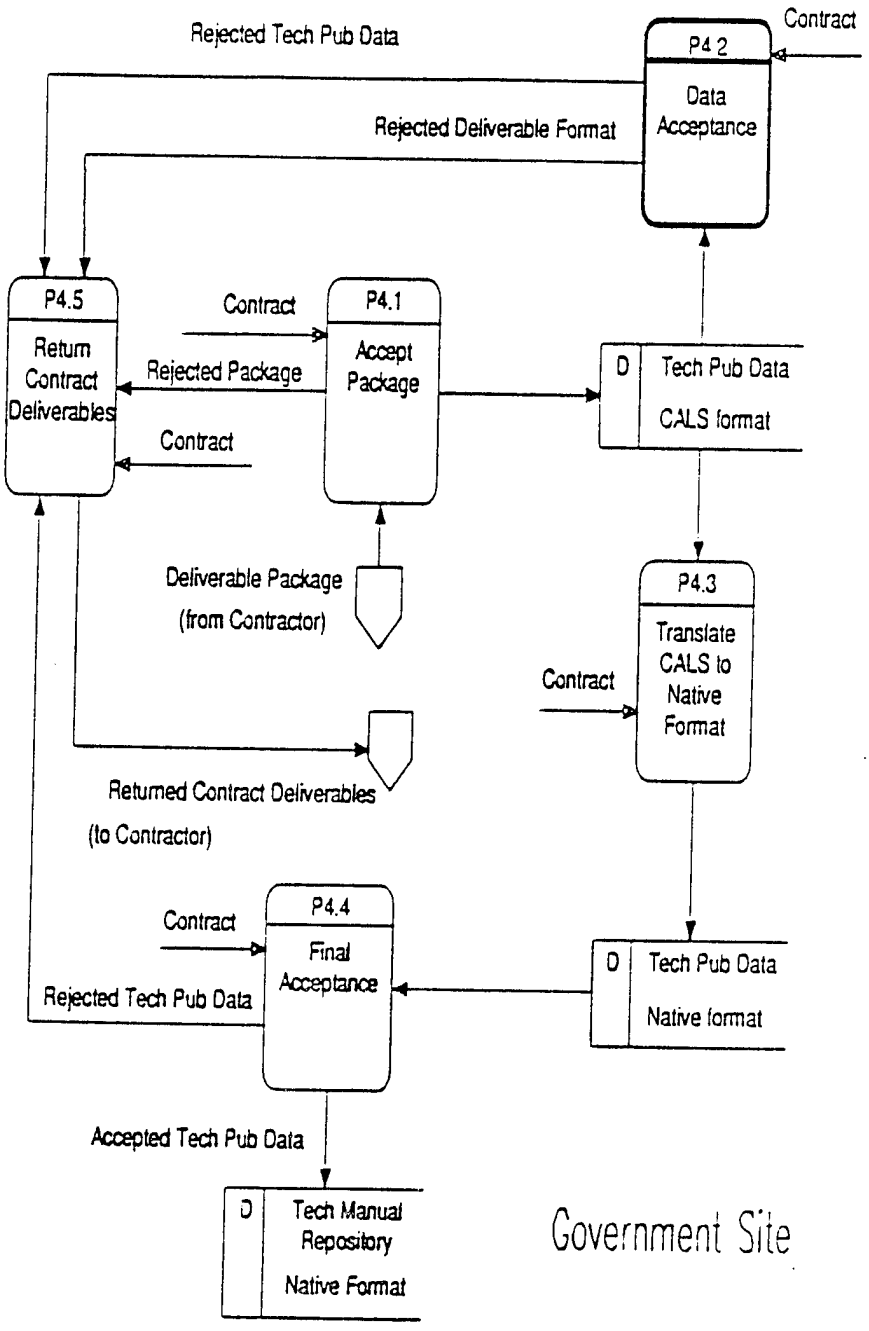


Figure 6 - Acceptance

Yet another alternative is to finalize the final output by an iterative process involving the contractor's review of the output produced by the government agency.

When the data passes the final acceptance, it will be stored in the repository. Any contract specified documentation will be completed.

Return Contract Deliverables (P4.5). The procurement should have provisions for the return and correction of digital data which is not accepted. The contract should specify what contract deliverables are returned and under what circumstances they can be returned. For example, in the case improper shipping and/or damage during shipping, the unopened package should be returned to the contractor. In the case of an improperly formatted file, the volume containing the file may be returned. Data which is returned to the contractor is packaged and shipped in accordance with the terms of the contract.

9.0 DATA ACCEPTANCE

Data Acceptance is the application of digital data quality criteria to the deliverable files. This a part of the overall TM Acceptance Procedure which is described in Section 8.6 on page 25. It is a subprocess (P4.2) of the process shown in Figure 6. Data Acceptance is summarized in Figure 7.

Data Acceptance will uncover major anomalies in the basic areas of data format, text tagging, illustration image quality, raster page image quality, etc. These are anomalies which would cause later problems in areas such as translation, producing output, indexing. When such problems are found, the data can be rejected before it is translated or stored in a repository. Uncovering such anomalies in the Technical Manual data while it is still in deliverable form will be more economical than uncovering them while trying to produce a formatted output or after the data is stored in a repository.

Data acceptance uses computer-assisted automatic quality assurance techniques to evaluate large quantities of data with little or no human intervention. Some of these techniques are: raster image quality evaluation, IGES entity evaluation, SGML syntax check, file cross-reference check, text spell check, and text grammar check.

Since Data Acceptance deals with deliverable files in exchange format, it does not address the content or presentation format of the document. It assumes that the content has been reviewed and approved. It assumes that the presentation format will be reviewed and approved in the Final Acceptance process described in Section 8.6 on page 25.

Format Verification (P4.2.1). The deliverable is subjected to a data format check using the criteria described in Section 7.1 on page 14.

Raster Page Document Acceptance (P4.2.2). Each Raster Page Image Document of a deliverable is subjected to Raster Page Image Document Acceptance. Raster Page Image Document Acceptance is described in more detail in Section 9.1 on page 30.

SGML Tagged Document Acceptance (P4.2.3). Each SGML Tagged Document of a deliverable is subjected to SGML Tagged Document Acceptance. SGML Tagged Document Acceptance is described in more detail in Section 9.2 on page 32.

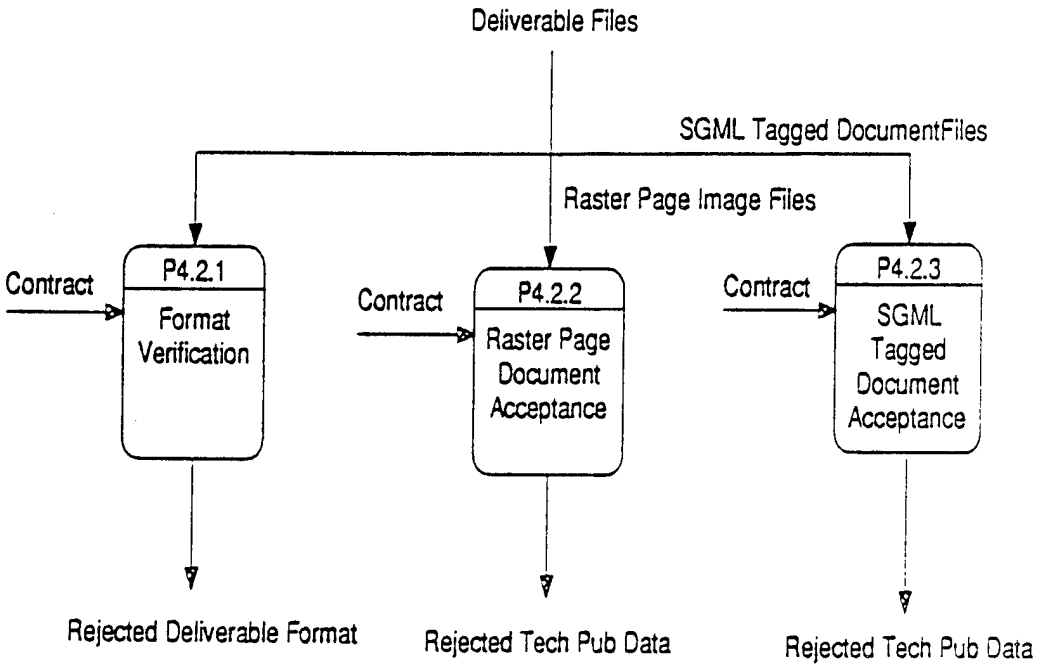


Figure 7 - Data Acceptance

9.1 Raster Page Image Acceptance

Raster Page Image Acceptance is illustrated in Figure 8. Raster Page Image documents are digital data which have been converted from existing hardcopy or microfiche documents. A Raster Page Image document is subjected to the following data acceptance checks. A document which fails any of these checks is rejected and returned to the contractor.

Image Quality Verification (P4.2.2.1). Each file, or page, of a raster page document is subjected to a raster image quality check using the criteria described in Section 7.3 on page 15. The image quality of each page is checked for acceptable density, noise, and verticality to ensure that it is legible and reproducible.

Header Data Verification (P4.2.2.2). Each file, or page, of a raster page document is subjected to an identification data quality check using the criteria described in Section 7.4 on page 16. The header data is checked to ensure that each individual file can be correctly indexed. Each file is checked for matching document number in the 'dstdocid:' record and the 'dstdocid:' record in the declaration file. Each file is checked for matching page number in the 'txtfilid:' header record and page number in the raster image. The entire document is checked for missing or duplicate pages.

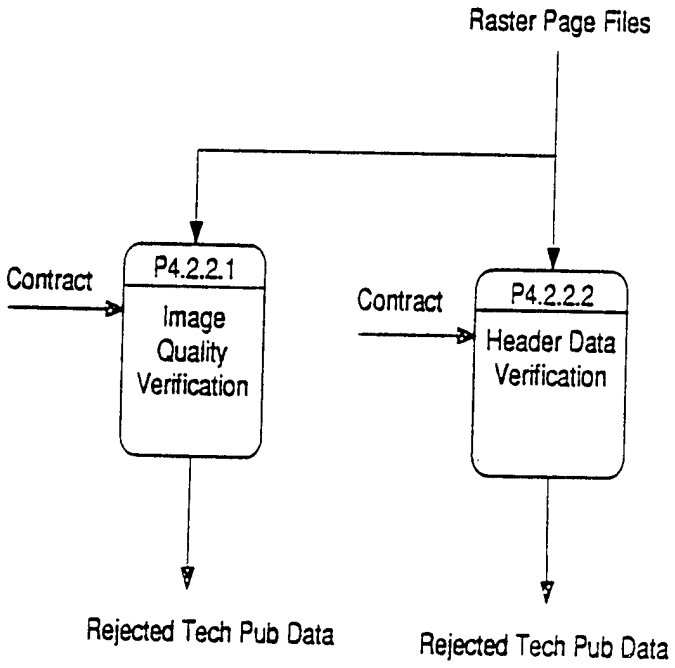


Figure 8 - Raster Page Document Acceptance

9.2 SGML Document Acceptance

SGML Document Acceptance is shown in Figure 9. SGML tagged text documents may be new Technical Manual procurement or conversion of existing tech manuals. An SGML document is subjected to the following data acceptance checks. A document which fails any of these checks is rejected and returned to the contractor.

Header Data Verification (P4.2.3.1). Each file of an SGML tagged document is subjected to an identification data quality check using the criteria described in Section 7.4 on page 16.

Text Accuracy Verification (P4.2.3.2). If a document has text which was generated by character recognition, the text files are subjected to spell checking and grammar checking using the character recognition accuracy criteria described in Section 7.5 on page 18.

SGML Syntax Verification (P4.2.3.3). Each text file of an SGML tagged document is subjected to an SGML syntax check using the criteria described in Section 7.6 on page 18. A conforming document must reference the appropriate Document Declaration Set using the appropriate public identifier.

DTD Syntax Verification (P4.2.3.4). The DTD file of an SGML tagged document is subjected to a DTD syntax check using the criteria described in Section 7.7 on page 18.

FOSI Syntax Verification (P4.2.3.5). A FOSI file of an SGML tagged document is subjected to a FOSI syntax check using the criteria described in Section 7.8 on page 18.

Raster Illustration File (Check) Quality Verification (P4.2.3.8). Each raster illustration file of an SGML tagged document is subjected to a format check as described in Section 7.1 on page 14. It is also subjected to a raster image quality check using the criteria described in Section 7.3 on page 15. The image quality of each illustration is checked for acceptable contrast, noise, and verticality to ensure that it is legible and reproducible.

IGES Illustration File (Check) Quality Verification (P4.2.3.7). Each IGES illustration file of an SGML tagged document is subjected to a format check as described in Section 7.1 on page 14.

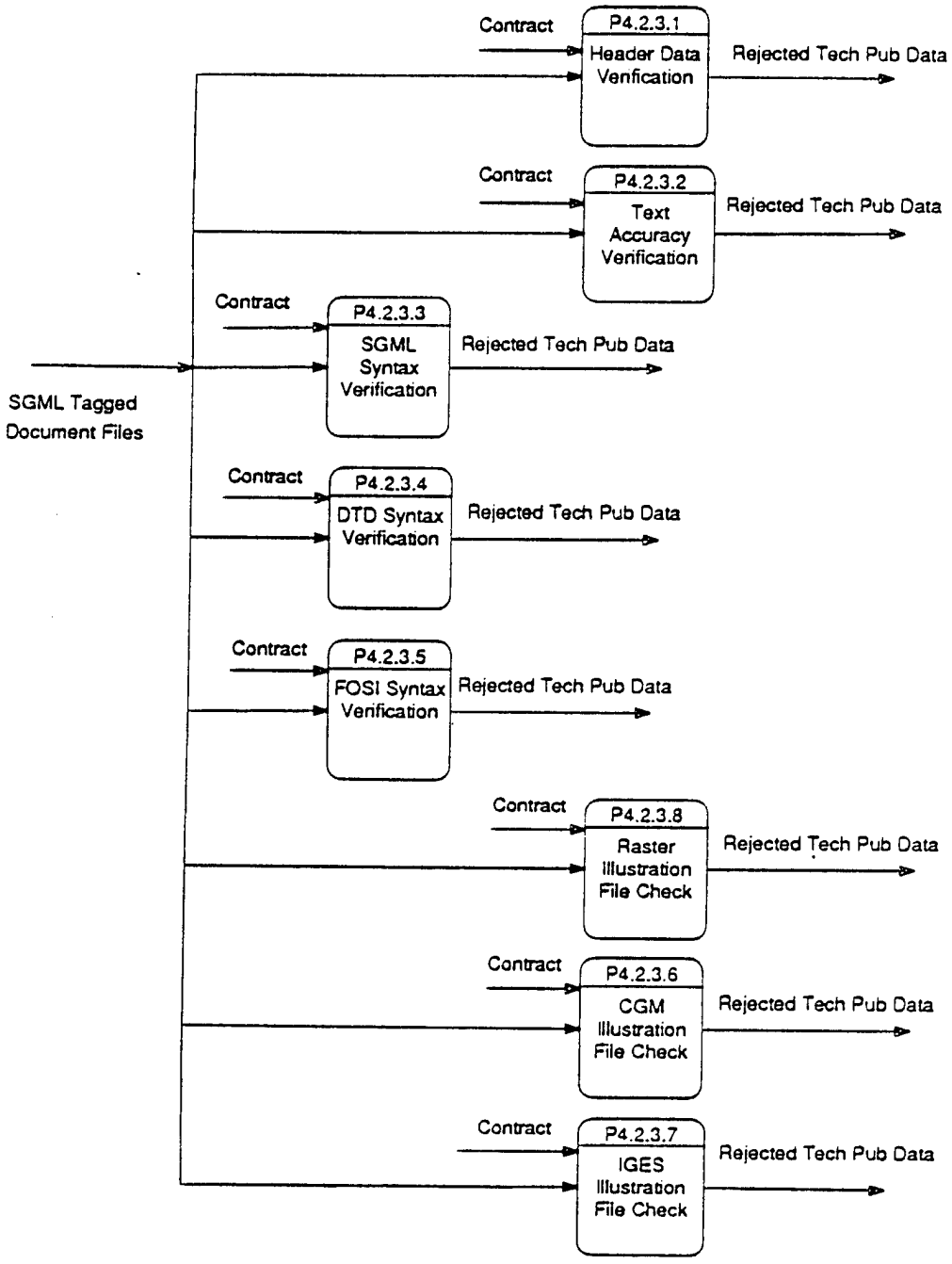


Figure 9 - SGML Tagged Document Acceptance

CGM Illustration File (Check) Quality Verification (P4.2.3.6). Each CGM illustration file of an SGML tagged document is subjected to a format check as described in Section 7.1 on page 14.

10.0 CONTRACT CONSIDERATIONS

The contract is the guiding document for the implementation of quality assurance and data acceptance. It determines how, when, where, and by whom quality assurance and data acceptance procedures are implemented. The contract must provide specific guidance for resolving important issues. This section identifies some of these issues and the options available for resolving each.

Data Format. The contract will specify that the data will be stored as SGML tagged text or raster page image files. Raster page image data consists of raster images compressed in CCITT Group IV format.

Delivery Mode. The contract will specify that the mode of digital data delivery complies with the options identified in MIL-STD-1840A, *Automated Interchange of Technical Information.* Currently, 9-track magnetic tape and optical disk are specified as acceptable transfer media. Magnetic tape formatting will be in accordance with MIL-STD-1840A. Optical disk formatting will be in accordance with contract specifications or future releases of government standards. Delivery by telecommunications may also be specified in a future release.

Computer Assisted Data Acceptance Procedures. The contract will specify that computer assisted data acceptance procedures be applied to the deliverable digital data during the acceptance function of data procurement.

Location and Schedule of Data Acceptance. The contract will specify the location and schedule of data acceptance. Acceptance may take place at a repository or other government site.

Data Acceptance Personnel. The contract will specify who is to perform data acceptance. A memorandum of understanding between the procuring agency, the repository and other concerned agencies will be generated to identify areas of responsibility for each.

Facility Requirements. The contract will specify what facilities are required to perform data acceptance. This will include, as a minimum: space, utilities, computer hardware and software, peripherals and interfaces. The provider of these facilities will be identified. The contract will state that these facilities will be available to the government inspector during data acceptance.

Acceptance Criteria. The contract will specify the acceptance criteria that will be used to judge the deliverable data. This will include the definition of acceptable data format, image quality and image identification data; inspection methods and quantitative acceptance criteria.

Disposition Determination. The contract will specify the action to be taken when the data is rejected or accepted as a result of data acceptance. This will include the destination of the deliverable and the documents that will accompany the deliverable data.

Data Certification. The contract will state that the contractor will inspect each data file and will produce a document certifying that the inspections have taken place.

Data Warranties. The contract will specify the terms and conditions of the warranty of deliverable data. This will include notification procedures and return/correction procedures. The contract will specify the time period during which rejected deliverables can be returned.

Source System Validation. The contract will specify that the contractor has completed or will complete a validation process to demonstrate the ability to produce digital data in accordance with standards. The contract will also specify that the contractor has a government sanctioned quality assurance program in place for the production and quality assurance of digital data.

Transfer Media Format Validation. The contract will specify that the contractor has completed or will complete a validation process to demonstrate the ability to produce digital data on transfer media in accordance with MIL-STD-1840A.

APPENDIX A - GLOSSARY

ASCII	American Standard Code for Information Interchange
CAD	Computer Aided Design
CADA	Computer Assisted Data Acceptance
CAGE	Commercial and Government Entity
CALS	Computer-aided Acquisition and Logistic Support
CCITT	Comité Consultatif Internationale de Télégraphique et Téléphonique (English translation: International Consultative Committee on Telegraphy and Telephony)
CDRL	Contract Data Requirements List
CGM	Computer Graphics Metafile
CPU	Central Processing Unit
CTN	CALS Test Network
DA	Data Acceptance
DID	Data Item Description
DoD	Department of Defense
DSREDS	Digital Storage and Retrieval Engineering Data System
DTD	Document Type Definition
EDCARS	Engineering Data Computer Assisted Retrieval System
EDMICS	Engineering Data Management Information and Control System
ER	Entity Relationship
EO	Engineering Order
FOSI	Formatting Output Specification Instance
IGES	Initial Graphics Exchange Specification
I/O	Input/Output
PDL	Page Description Language
PM CALS	Project Manager CALS
QA	Quality Assurance
RAM	Random Access Memory

SGML	Standard Generalized Markup Language
SOW	Statement of Work
TM	Technical Manual
TO	Technical Order
TP	Technical Publication

APPENDIX B - REFERENCE DOCUMENTS

MIL-D-28000	Digital Representation for Communications of Product Data: IGES Application Subsets
MIL-M-28001	Markup Requirements and Generic Style Specification for Electronic Printed Output and Exchange of Text
MIL-R-28002A	Requirements for Raster Graphics Representation in Binary Format
MIL-D-28003	Digital Representation for Communications of Product Data: CGM Application Subsets
MIL-M-38784	Military Specification Manuals, Technical: General Style and Format Requirements
MIL-STD-1840A	Automated Interchange of Technical Information.
	Contractor Site Digital Data Acceptance/Quality Assurance Procedures, 9 February 1990, Department of the Army, PM CALS
	DSREDS/EDCARS Site Digital Data Acceptance/Quality Assurance Procedures, 9 February 1990, Department of the Army, PM CALS
	Model - Engineering Data, 9 December 1990, Department of the Army, PM CALS
	Computer Assisted Data Acceptance Procedures, 22 February 1991, Department of the Army, PM CALS

This page was intentionally left blank.